Chapter 6
Exploring Data: Relationships

For All Practical Purposes: Effective Teaching

- A characteristic of an effective instructor is fairness and consistency in grading and evaluating student performance.
- A sense of humor is appreciated by students. They also appreciate knowing you are approachable and enthusiastic. Once your organization and attitude are known to students, achieving interaction with students that is positive takes a relatively small amount of effort on your part.

Chapter Briefing

In this chapter, you will be mainly examining relationships between two variables. This will be done initially by examining *scatterplots*, which consist of points in the plane. When a scatterplot suggests a linear relationship between the variables, a regression line is often used to predict the value of y for a given value of x. In this chapter, you will interpret the results of a regression line as well as find the *least-squares regression line*. Since the least-squares regression line exists for any set of data, even one that does not follow a linear pattern, *correlation* will be examined as well as calculated. The correlation, r, is always a number between -1 and 1. Values of r close to 0 indicate a very weak linear relationship. Additionally, even though there is a strong correlation between two variables (either close to 1 or -1), the fact that this may not imply causation will be investigated. Finally, the effects of outliers on the least-squares regression line will be considered in this chapter.

You should find that students may need some additional background preparation in plotting points, graphing lines, understanding independent versus dependent variables, interpreting slope of a line as well as topics from Chapter 5 such as finding mean and standard deviation.

Being well prepared for class discussions with short examples and the knowledge that students may be lacking in the basics of graphing and interpreting lines and related topics are essential in order to help students focus on the main topics presented in this chapter. In order to facilitate your preparation, the **Chapter Topics to the Point** has been broken down into **Variable Types**, **Scatterplots**, **Regression Lines**, **Correlation**, **Least-Squares Regression**, **Effect of Outliers on Correlation and Regression**, and **Association Does Not Imply Causation**. Examples with solutions for these topics that do not appear in the text nor study guide are included in the *Teaching Guide*. You should feel free to use these examples in class, if needed.

Since you may be asked to demonstrate the techniques of this chapter using graphing calculator, the *Teaching Guide* includes the feature **Teaching the Calculator**. It includes brief calculator instructions with screen shots from a TI-83.

The last section of this chapter of *The Teaching Guide for the First-Time Instructor* is **Solutions to Student Study Guide ✎ Questions**. These are the complete solutions to the three questions included in the *Student Study Guide*. Students only have the answers to these questions, not the solutions.

Chapter Topics to the Point

Variable Types

We will be using data sets that have two types of variables. A **response variable** measures an outcome or result of a study. An **explanatory variable** is a variable that we think explains or causes changes in the response variable. Typically, we think of the explanatory variable as x (the independent variable) and the response variable as y (the dependent variable).

Teaching Tip

When the terminology of explanatory and response variables is discussed, relate them to input versus output. You may choose to avoid using the word "function" and instead give real-world types of examples.

Teaching Tip

A good place to start class discussion of variable types is to ask students about things in their everyday life that they feel have a relationship. Then further ask if they believe one *explains* the other or if they believe one is in *response* to the other.

Example

For the following situations, is it more reasonable to simply explore a relationship between the two variables or to view one variable as the explanatory variable and the other as the response? If one variable should be the explanatory variable, x, and the other as the response, y, state which variable should be considered as which.

a) The number of words on page and the thickness of the book
b) Average of parents' height and the height of their child
c) Shoe size and the age of a child

Solution

a) relationship only
b) The average of parents' height would be the explanatory variable and the height of their child would be the response variable.
c) The shoe size would be the response variable and the age of the child would be the explanatory variable.

Scatterplots

Graphs are useful for recognizing connections between two variables. By taking the data as a set of ordered pairs (x, y), points are plotted in a plane to form a **scatterplot**. The scatterplot is the simplest graphical representation, showing the relationship between an explanatory variable (on the horizontal axis) and a response variable (on the vertical axis). In general, when we create a scatterplot we are interested in three things in the **overall pattern**. The pattern can be described by the following.

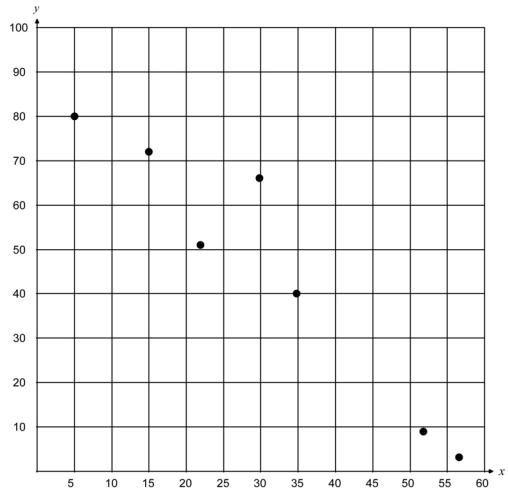
- *form*: straight – line, for example
- *direction*: positive association or negative association (slope of a line)
- *strength*: A stronger relationship would yield points quite close to the line, a weaker one would have more points scattered around the line.

Example

Draw a scatterplot showing the relationship between the observed variables x and y, with the data given in the table below. Make some observations of the overall pattern in the scatterplot.

x	5	15	22	30	35	52	57
y	80	72	51	66	40	9	3

Solution



It appears that the points indicate a strong linear relationship with a negative association.

Teaching Tip

You may chose to tell the students that when they are constructing a scatterplot, they should scan the values of the explanatory and the response variables in order to get a feeling for both the range and scale on the x- and y- axes.

Regression Lines

A straight line drawn through the heart of the data and representing a trend is called a **regression line**, and can be used to predict values of the response variable. The equation of a regression line will be $\boldsymbol{y = a + bx}$, where a is the y-intercept and b is the slope of the line.

Teaching Tip

You may choose to emphasize to students that the form of the regression line used is $y = a + bx$. Students may not pay attention to this point and think of the equation of a line of this form as $y = ax + b$ or $y = mx + b$. In doing so, students may confuse the role of b. In the case of the regression line, b is the slope, not the y-intercept.

Teaching Tip

You may choose to review simple examples of graphing of the form $y = a + bx$ such as $y = 2 + 3x$ and $y = 4 - 0.5x$. In doing so, review the role of the y-intercept, slope, and if applicable, the x-intercept.

Example

Suppose that the slope of the regression line of car value for a certain model of car is $b = -1966$ when we measure time x in years ($x = 0$ indicates present day) and value y in dollars. Also, suppose the y-intercept is 32,500.

a) State the equation of the regression line.

b) Give a meaning to the slope and the y-intercept.

c) Graph the regression line stated in Part a and use the "up and over" method described in Section 6.2 to determine the approximate value of the car after 10 years.

d) Using the graph, when will the car have no value?

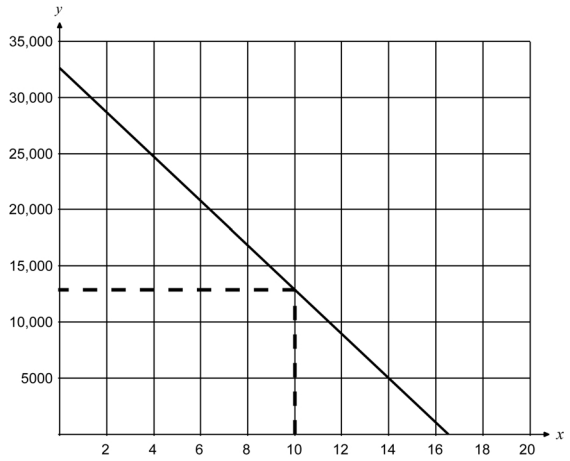
e) Verify your approximations found in Parts c and d by using the equation of the regression line.

Solution

a) $y = 32,500 - 1966x$

b) The slope represents the number of dollars the value of the car decreases in one year. The y-intercept could represent the price paid for the car. It could also represent the value of the car immediately when it is purchases as a resale value. Typically, when a new car is purchased, it immediately loses value so the amount paid may not be the value of the car immediately after purchase.

c) It appears that the approximate value of the car after 10 years is \$12,800.



d) After the 16th year or during the 17th year, the car no longer has a value.

e) For Part c, we want to know what is y when $x = 10$.

$$y = 32,500 - 1966(10) = 32,500 - 19,660 = \$12,840$$

For Part d, we want to know what is x when $y = 0$.

$$0 = 32,500 - 1966x$$

$$1966x = 32,500$$

$$x = \frac{32,500}{1966} \approx 16.53 \text{ years}$$

Correlation

The following is the **formula for correlation** given the means and standard deviations of the two variable x and y for the n individuals.

$$r = \frac{1}{n-1}\left[\left(\frac{x_1-\bar{x}}{s_x}\right)\left(\frac{y_1-\bar{y}}{s_y}\right)+\left(\frac{x_2-\bar{x}}{s_x}\right)\left(\frac{y_2-\bar{y}}{s_y}\right)+....+\left(\frac{x_n-\bar{x}}{s_x}\right)\left(\frac{y_n-\bar{y}}{s_y}\right)\right]$$

The correlation, r, measures the strength of the linear relationship between two quantitative variables; r always lies between –1 and 1. **Positive *r*** means the quantities tend to increase or decrease together; **negative *r*** means they tend to change in opposite directions, one going up while the other goes down. If ***r* is close to 0**, that means the quantities are fairly independent of each other.

Example

Consider the following small data set.

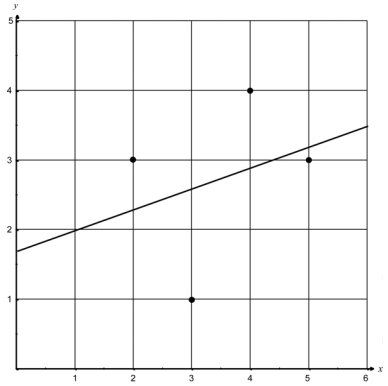
x	2	3	4	5
y	3	1	4	3

a) Make a scatterplot and with a general guideline that there should be roughly the same number of points above a regression as below, sketch a possible regression line and state if the slope of the regression line (and then in turn the correlation) should be negative or positive.

b) Calculate the correlation using the formula.

Solution

a) The actual regression line has been drawn below. The slope of the regression line is positive.



b)

i	Observations x_i	Observations y_i	Deviations $x_i-\bar{x}$	Deviations $y_i-\bar{y}$	Squared deviations $(x_i-\bar{x})^2$	Squared deviations $(y_i-\bar{y})^2$
1	2	3	– 1.5	0.25	2.25	0.0625
2	3	1	– 0.5	– 1.75	0.25	3.0625
3	4	4	0.5	1.25	0.25	1.5625
4	5	3	1.5	0.25	2.25	0.0625
sum	**14**	**11**	**0**	**0**	**5**	**4.75**

$\bar{x} = \frac{14}{4} = 3.5$, $\bar{y} = \frac{11}{4} = 2.75$, $s_x^2 = \frac{5}{4-1} = \frac{5}{3} \approx 1.66667$, $s_x \approx \sqrt{1.66667} \approx 1.2910$,

$s_y^2 = \frac{4.75}{4-1} = \frac{4.75}{3} \approx 1.58333$, and $s_y \approx \sqrt{1.58333} \approx 1.2583$.

Continued on next page

Since

$$r = \frac{1}{n-1}\left[\left(\frac{x_1-\bar{x}}{s_x}\right)\left(\frac{y_1-\bar{y}}{s_y}\right)+\left(\frac{x_2-\bar{x}}{s_x}\right)\left(\frac{y_2-\bar{y}}{s_y}\right)+\left(\frac{x_3-\bar{x}}{s_x}\right)\left(\frac{y_3-\bar{y}}{s_y}\right)+\left(\frac{x_4-\bar{x}}{s_x}\right)\left(\frac{y_4-\bar{y}}{s_y}\right)\right],$$

we have the following.

$$r \approx \frac{1}{4-1}\left[\left(\frac{-1.5}{1.2910}\right)\left(\frac{0.25}{1.2583}\right)+\left(\frac{-0.5}{1.2910}\right)\left(\frac{-1.75}{1.2583}\right)+\left(\frac{0.5}{1.2910}\right)\left(\frac{1.25}{1.2583}\right)+\left(\frac{1.5}{1.2910}\right)\left(\frac{0.25}{1.2583}\right)\right]$$

$$= \frac{1}{3}\left[\frac{-0.375}{1.6244653}+\frac{0.875}{1.6244653}+\frac{0.625}{1.6244653}+\frac{0.375}{1.6244653}\right]$$

$$= \frac{1}{3}\left[\frac{1.5}{1.6244653}\right] = \frac{0.5}{1.6244653} \approx 0.3078$$

Teaching Tip

If a linear regression line is easily visualized, you may choose to tell students that a line with a positive slope will indicate positive correlation and a negative slope will indicate negative correlation. How close a correlation is to either –1 (from above) or 1 (from below) indicates the strength of correlation – that is, how much the scatterplot resembles a line. Strong versus weak correlation is in the context of the results. Some arbitrary, but acceptable limits for $|r|$, would be as follows.

$0.00 - 0.19$ very weak
$0.20 - 0.39$ weak
$0.40 - 0.59$ moderate
$0.60 - 0.79$ strong
$0.80 - 1.00$ very strong

Teaching Tip

It should be emphasized that correlation does not have units. Realizing that standard deviation and data have the same units, you can see why correlation does not have units by looking at the summation formula for correlation, $r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$. Naturally, you would want to use the expanded form of this formula when explaining why correlation has no units. Also, correlation will not change when measurement units change.

Example

Consider the data from an earlier example.

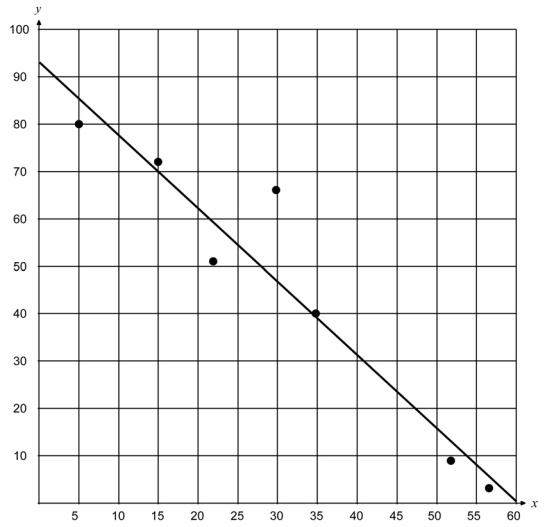
x	5	15	22	30	35	52	57
y	80	72	51	66	40	9	3

a) Draw the regression line $y = 93.1 - 1.53x$ on the scatterplot.

b) Choose a category of correlation (very weak, weak, moderate, strong, or very strong) based on how well the line fits the data.

Solution

a)



b) It appears that the correlation should be very strong. In fact the correlation is $r \approx -0.9548$.

Least-Squares Regression

The **least-squares regression line** runs through a scatterplot of data so as to be the line that makes the sum of the squares of the vertical deviations from the data points to the line as small as possible. This is often thought of as the "line of best fit" to the data. When we do least-squares regression, we are looking for the line that fits the data best. In finding this line, we are trying to minimize the error and find the line that differs least from all the data points. Since the linear equation is a formula to predict the value of y given a value of x, it should make sense that we are trying to minimize the error in estimating all the y values (vertical line segments).

The formula for the equation of the least-square regression line for a data set on an explanatory variable x and a response variable y depends on knowing the means of x and y, the standard deviations of x and y, and their correlation r. It produces the slope and intercept of the regression line. The least-square regression line is predicted $y = a + bx$, where $b = r\frac{s_y}{s_x}$ (slope) and $a = \overline{y} - b\overline{x}$.

Teaching Tip

If you are using a graphing calculator, many of them will calculate the least-squares regression line in two ways. One is $y = ax + b$ and the other is $y = a + bx$. The calculated correlation will be the same. The only difference in the exchange is the roles of a and b as slope and y-intercept. Also, make sure your students are clear as to the course expectations of using technology in doing such calculations.

Example

Consider the data from an earlier example.

x	2	3	4	5
y	3	1	4	3

Find the equation of the least-squares regression line.

Solution

From our earlier example, we have $\bar{x} = 3.5$, $\bar{y} = 2.75$, $s_x \approx 1.2910$, $s_y \approx 1.2583$, and $r \approx 0.3078$. Since $b = r\frac{s_y}{s_x} = 0.3078\left(\frac{1.2583}{1.2910}\right) \approx 0.300$ and $a = \bar{y} - b\bar{x} = 2.75 - 0.300(3.5) = 1.7$, the least-square regression line is $y = a + bx = 1.7 + 0.3x$.

Effect of Outliers on Correlation and Regression

Both the correlation r and the least-squares regression line can be strongly influenced by a few outlying points. Never trust a correlation until you have plotted the data.

Example

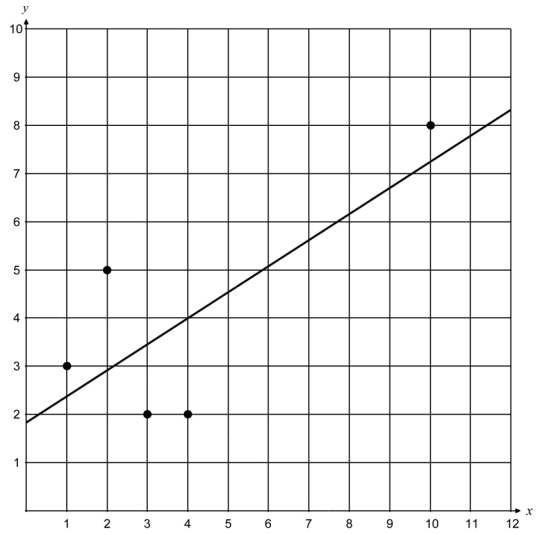
Consider the following data.

x	1	2	3	4	10
y	3	5	2	2	8

a) Draw a scatterplot and use a calculator to determine the correlation and least-squares regression line.
b) Remove the outlier from the data and compute again the correlation and least-squares regression line. Comment on the results.

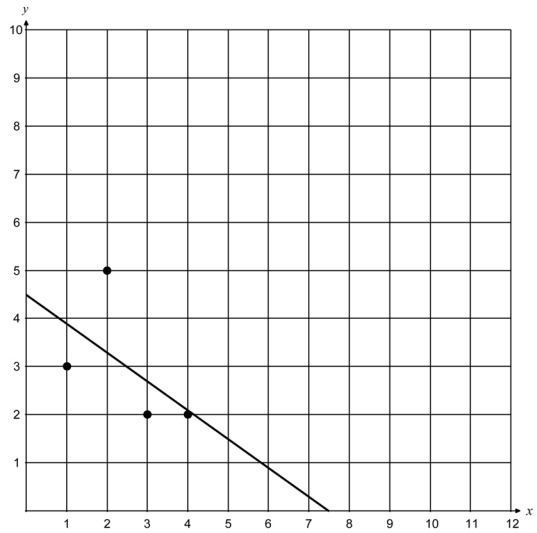
Solution

a) The correlation, r, is approximately 0.749. The equation of the least-squares regression line is $y = 1.84 + 0.54x$. The data points are plotted below along with the graph of the least-squares regression line.



Continued on next page

b) The correlation, r, is approximately -0.548. The equation of the least-squares regression line is $y = 4.5 - 0.6x$. The data points are plotted below along with the graph of the least-squares regression line.



With the outlier, there was a strong positive correlation. Without the outlier, the correlation is weaker and negative.

Teaching Tip

Question 3 in the *Student Study Guide* yields the outcome that when the outlier is removed, a strong correlation occurs. The complete solution to the exercise appears at the end of this chapter. You may wish to make the point that an outlier may cause either a higher or a weaker correlation.

Association Does Not Imply Causation

Correlation and regression *describe* relationships. *Interpreting* relationships requires more thought. Try to think about the effects of other variables prior to drawing conclusions when interpreting the results of correlation and regression. An association between variables is not itself good evidence that a change in one variable actually causes a change in the other!

Teaching Tip

You may choose to give oddball relations that relate to the stock market. For example, in January of 2001, the economy was worse than in January of 2000. The following relation was made.

- As the sale of corrugated boxes fell, so did the economy.
    - January 2000: Sales were 33.220 billion square feet; January 2001: Sales were 33.104 billion square feet

There are other interesting relations to start classroom discussions that demonstrates association does not imply causation. These include the relationships between the stock market and AFC teams winning the Super Bowl or hemline length of skirts.

Teaching Tip

Make sure you are aware of the requirements placed on students regarding how technology should be used in these calculations. If the faculty wishes to integrate forms of technology such as graphing calculators with statistical capabilities or spreadsheets, make sure you can demonstrate their use in the classroom.

Teaching the Calculator

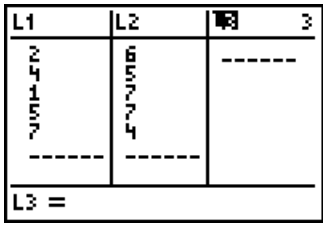
Example 1

Create a scatterplot given the following data.

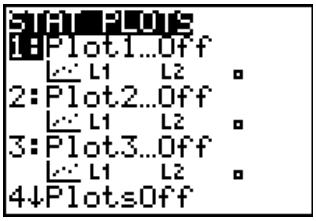
x	2	4	1	5	7
y	6	5	7	7	4

Solution

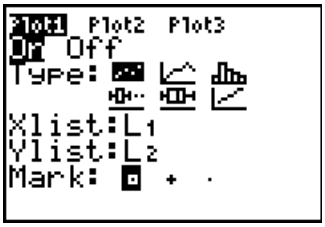
First enter the data as described in Chapter 5 section of Learning the Calculator. You should have the following screen.



In order to display a scatterplot, you press [2nd] then [Y=]. This is equivalent to [STAT PLOT]. The following screen (or similar) will appear.

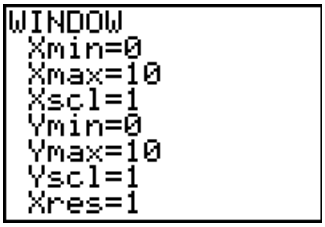


You will need to turn a stat plot On and choose the scatterplot option ( ). You will also need to make sure Xlist and Ylist reference the correct data. In this case L1 and L2, respectively.

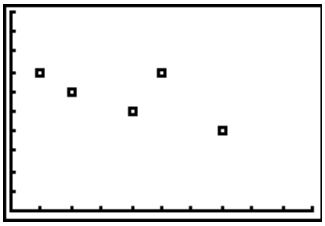


As was noted in the Chapter 5 section of Learning the Calculator, you will need to make sure that no other graphs appear on your scatterplot.

You will next need to choose an appropriate window. By pressing [WINDOW] you need to enter an appropriate window that includes your smallest and largest pieces of data in L1. These values dictate your choices of Xmin and Xmax. You will also need to enter an appropriate window that includes your smallest and largest pieces of data in L2. These values dictate your choices of Ymin and Ymax. Choose convenient values for Xscl and Yscl. In this case, 1 for each would be convenient.



Next, we display the histogram by pressing the [GRAPH] button.



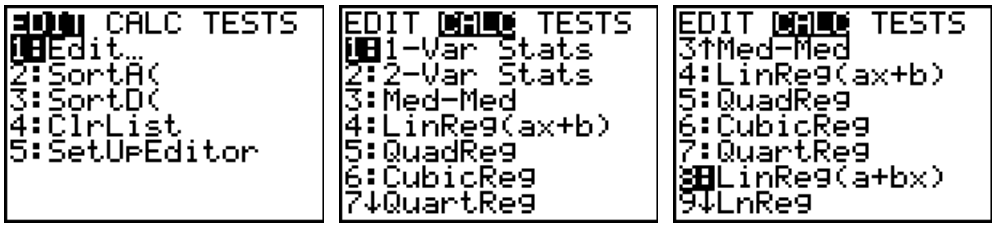
Example 2

Find and graph the least-squares regression line for the following data.

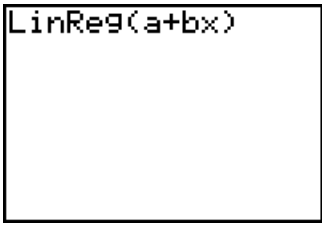
x	2	4	1	5	7
y	6	5	7	7	4

Solution

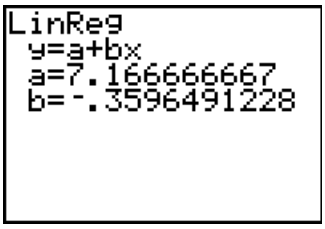
With data already entered, press the [STAT] button. Toggle to the right for CALC. Toggle down to 8:LinReg(a+bx) and press [ENTER] .



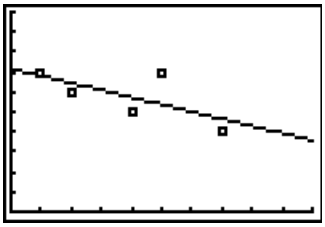
Instead of toggling down to 8:LinReg(a+bx) and pressing [ENTER] , you could alternatively press the 8 button ([8]). In either case the following screen will appear.



By pressing [ENTER] , you may get the following screen. Your screen may have more information.

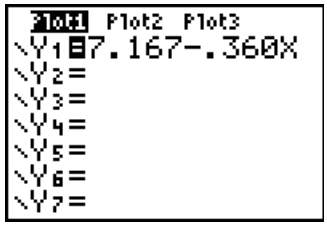


There are several ways to obtain the following graph of the least-squares line along with the scatterplot.



In all three methods, you will need to press [Y=] in order to enter the equation.

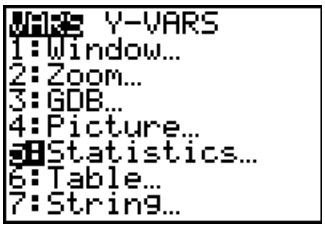
Method I: Type in the equation of the regression line, $y = a + bx$, by rounding the values of a and b.

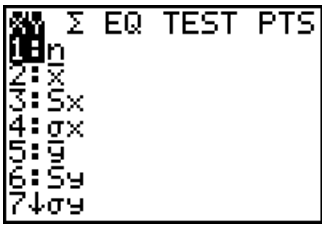


Press [GRAPH] in order to obtain the graph. This is the easiest method.

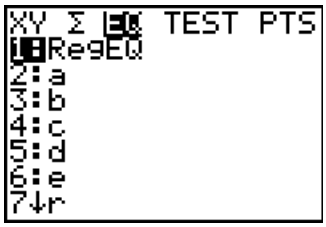
Method II: Place the equation of the regression line, $y = a + bx$, up to the accuracy of the calculator.

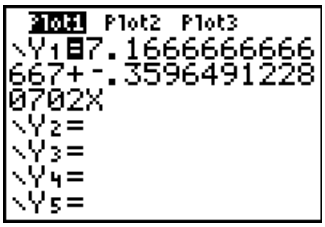
To do this, you press [VARS] then toggle down to 5:Statistics and press [ENTER]. You could alternatively press the 5 button ([5]).





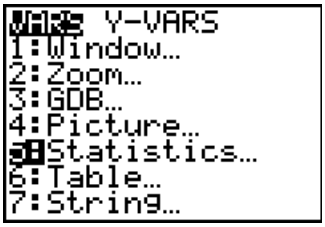
Toggle to the right to the EQ menu and press [ENTER].



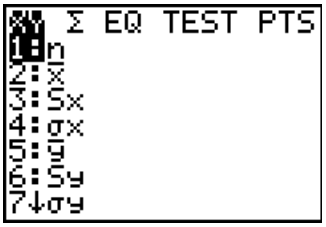


Press [GRAPH] in order to obtain the graph.

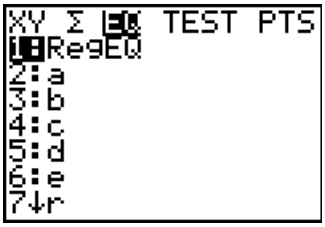
Method III: Place the equation of the regression line, $y = a + bx$, in general into your equation editor. To do this, you press [VARS].



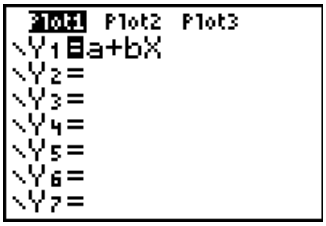
Toggle down to 5:Statistics and press [ENTER]. You could alternatively press the 5 button ([5]).



Toggle to the right to the EQ menu and toggle down to 2:a and press [ENTER]. You could alternatively press the 2 button ([2]). After pressing the plus button ([+]), press [VARS] again and run through the similar procedure to insert *b*.



Finally, press [X,T,Θ,n] to get the following screen.



Press [GRAPH] in order to obtain the graph. Although this is the hardest method, you only need to do it once. When you don't wish for the equation to be graphed, simply de-select the relation by the method described in Chapter 5 of Teaching the Calculator.

Example 3

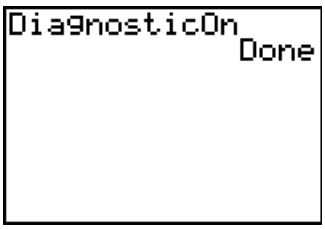
Find the correlation for the following data.

x	2	4	1	5	7
y	6	5	7	7	4

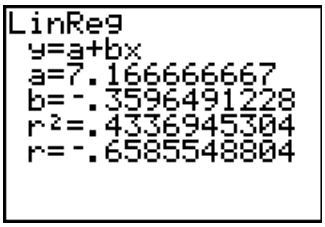
Solution

You may have already obtained the correlation when you obtained the least-squares line in Example 2. If not, you need to activate the DiagnosticOn feature. To do this, press [2nd] then [0]. This will take you to the CATALOG menu.

Toggle down to DiagnosticOn and press [ENTER] twice. You should get the following screen.



Finally, follow the instructions in Example 2 to obtain the least-squares line. The correlation is $r \approx -0.659$.



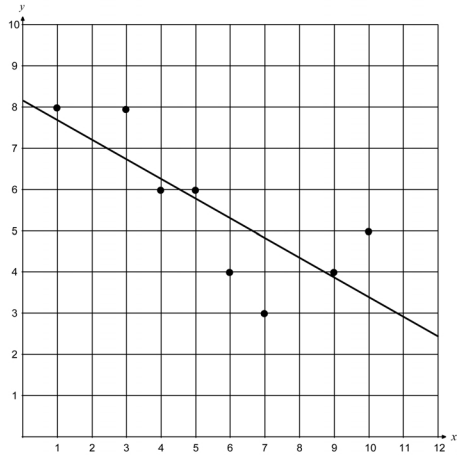
Solutions to Student Study Guide ✎ Questions

Question 1

Given the following data with regression line $y = 8.19 - 0.477x$ (obtained from a computer program). Determine which point is closest to the regression line and which point is furthest. Do this by making a scatterplot, drawing the regression line, and visually determining which point is closest and which point is furthest from the line.

x	10	1	6	7	4	9	3	5
y	5	8	4	3	6	4	8	6

Solution



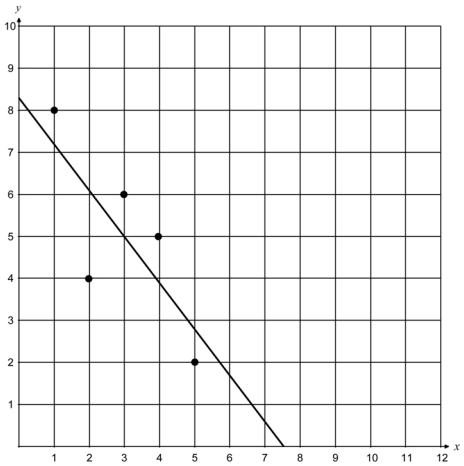
$(7,3)$ appears to be farthest away, and $(9,4)$ appears to be closest to the regression line.

Question 2

Given the following data, compute the correlation and least-squares regression line by hand.

x	1	2	3	4	5
y	8	4	6	5	2

Solution



Continued on next page

We have the following hand calculations.

i	Observations x_i	Observations y_i	Deviations $x_i-\bar{x}$	Deviations $y_i-\bar{y}$	Squared deviations $(x_i-\bar{x})^2$	Squared deviations $(y_i-\bar{y})^2$
1	1	8	– 2	3	4	9
2	2	4	– 1	– 1	1	1
3	3	6	0	1	0	1
4	4	5	1	0	1	0
5	5	2	2	– 3	4	9
sum	**15**	**25**	**0**	**0**	**10**	**20**

$\bar{x}=\frac{15}{5}=3$, $\bar{y}=\frac{25}{5}=5$, $s_x{}^2=\frac{10}{5-1}=\frac{10}{4}=2.5$, $s_x=\sqrt{2.5}\approx 1.5811$, $s_y{}^2=\frac{20}{5-1}=\frac{20}{4}=5$, and $s_y=\sqrt{5}\approx 2.2361$.

Since

$$r=\frac{1}{n-1}\left[\left(\frac{x_1-\bar{x}}{s_x}\right)\left(\frac{y_1-\bar{y}}{s_y}\right)+\left(\frac{x_2-\bar{x}}{s_x}\right)\left(\frac{y_2-\bar{y}}{s_y}\right)+\left(\frac{x_3-\bar{x}}{s_x}\right)\left(\frac{y_3-\bar{y}}{s_y}\right)+\left(\frac{x_4-\bar{x}}{s_x}\right)\left(\frac{y_4-\bar{y}}{s_y}\right)+\left(\frac{x_5-\bar{x}}{s_x}\right)\left(\frac{y_5-\bar{y}}{s_y}\right)\right],$$

we have the following.

$$\begin{aligned} r &\approx \frac{1}{5-1}\left[\left(\frac{-2}{1.5811}\right)\left(\frac{3}{2.2361}\right)+\left(\frac{-1}{1.5811}\right)\left(\frac{-1}{2.2361}\right)+\left(\frac{0}{1.5811}\right)\left(\frac{1}{2.2361}\right)+\left(\frac{1}{1.5811}\right)\left(\frac{0}{2.2361}\right)+\left(\frac{2}{1.5811}\right)\left(\frac{-3}{2.2361}\right)\right] \\ &= \frac{1}{4}\left[\frac{-6}{3.53549771}+\frac{1}{3.53549771}+\frac{0}{3.53549771}+\frac{0}{3.53549771}+\frac{-6}{3.53549771}\right] \\ &= \frac{1}{4}\left[\frac{-11}{3.53549771}\right]=\frac{-2.75}{3.53549771}\approx -0.7778 \end{aligned}$$

Since $b=r\frac{s_y}{s_x}=-0.7778\left(\frac{2.2361}{1.5811}\right)\approx -1.100$ and $a=\bar{y}-b\bar{x}=5-(-1.100)(3)=8.3$, we have the correlation, r, is approximately -0.778 and the least-square regression line is $y=a+bx=8.3-1.1x$.

Question 3

The following is data from a small company. The explanatory variable is the number of years with the company and the response variable is salary. Use a calculator to determine the correlation and least-squares regression line.

x	1 year	2 year	3 year	4 year	5 year
y	\$77,500	\$29,500	\$31,000	\$34,000	\$41,000

a) Using the regression line, project the salary of an employee that has been with the company 10 years. Comment on the results.

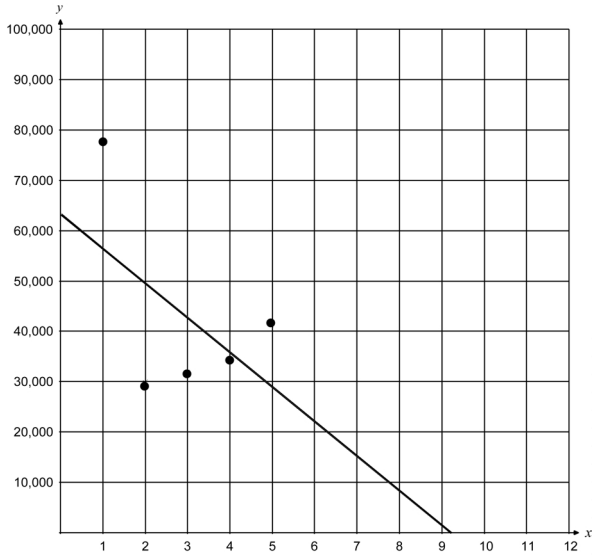
Remove the outlier from the data and compute again the correlation and least-squares regression line.

b) Using the regression line (without the outlier), project the salary of an employee that has been with the company 10 years. Comment on the results.

Solution on next page

Solution

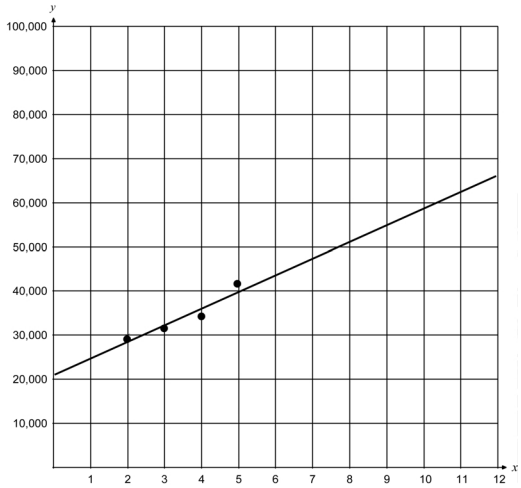
a) The correlation, r, is approximately -0.541. The equation of the least-squares regression line is $y = 63{,}150 - 6850x$. The data points are plotted below along with the graph of the least-squares regression line.



According to this regression line, it would be predicted that the salary of an employee that has been with the company for 10 years would be $63{,}150 - 6850(10) = 63{,}150 - 68{,}500 = -\$5350$.

It does not make sense for an employee to be earning a negative salary. Clearly, the one year employee represents an outlier. One possible explanation for this outlier is that he/she was hired one year prior as a president of this small company.

b) The correlation, r, is approximately 0.948. The equation of the least-squares regression line is $y = 20{,}750 + 3750x$. The data points are plotted below along with the graph of the least-squares regression line.



According to this regression line, it would be predicted that the salary of an employee that has been with the company for 10 years would be $20{,}750 + 3750(10) = 20{,}750 + 37{,}500 = \$58{,}250$.

This predicted salary seems reasonable. Notice that there is a high positive correlation between the number of years and the annual salary.